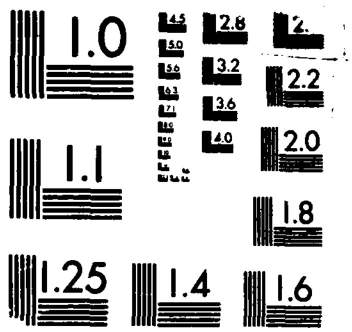
AD-A193 640 VISUAL MOTION PERCEPTION AND VISUAL ATTENTIVE PROCESSES 1/1
(U) NEW YORK UNIV N Y G SPERLING APR 88
AFOSR-TR88-0551 AFOSR-85-0364

UNCLASSIFIED F/G 6/4 NL



MICROCOPY RESOLUTION TEST CHART
URFAU OF STANDARDS-1963-A

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

DTIC FILE COPY

REPORT DOCUMENTATION PAGE

AD-A193 640

1a. REPORT SEC	1b. RESTRICTIVE MARKINGS
2a. SECURITY C	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFI	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 88-0551
6a. NAME OF PERFORMING ORGANIZATION New York University 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NL
6c. ADDRESS (City, State, and ZIP Code) 6 Washington Place, 8th Floor New York, NY 10003	7b. ADDRESS (City, State, and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-85-0364
8c. ADDRESS (City, State, and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F PROJECT NO. 2313 TASK NO. A5 WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Visual Motion Perception and Visual Attentive Processes

12. PERSONAL AUTHOR(S)

Dr. George Sperling

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 30 Sep 85 TO 30 Nov 87	Apr 88	16

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The motion projects described in this report include the specification of low level motion detection systems which are both Fourier and NonFourier in kind, investigations in visual persistence in motion, higher level issues in structure from motion, cue integration and decision theory, and object recognition. The attention projects concern the human ability to process information arriving simultaneously at different locations in the visual field and to coordinate concurrent visual and auditory inputs. A common goal of projects was the attention description of the human abilities and limitations in the allocation of mental processing resources, and correspondingly, the theoretical derivation of visual and auditory stimulus codes that take optimum advantage of human abilities.

DTIC ELECTE MAY 19 1988 S D E

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. John Tangney	22b. TELEPHONE (Include Area Code) (202) 767-5021 — 22c. OFFICE SYMBOL NL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

DTIC COPY INSPECTED 1

AFOSR-TR- 88-0551

Visual Motion Perception and Visual Attentive Processes

George Sperling, New York University

Grant AFOSR 85-0364

USAF Office of Scientific Research, Life Sciences Directorate

Final Progress Report, 1 October 1984 to 30 November 1987

Accession For
NTIS GRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

ABSTRACT

The main activities throughout this grant have been carrying out the experimental research set forth in the proposals (1980,1984), following up promising leads that developed in the course of this work, and preparing manuscripts for publication. The work is best described by the publications; these are appended. A brief overview is provided below.

FACILITIES

During the last granting period, a highly versatile laboratory has been created for research in almost any area of vision or cognition. NYU has provided generous space and renovations for better air cooling and electrical connections. The Laboratory is organized around a PDP VAX 11/750 with a UNIX operating system; everyone has a terminal and is connected to it. The main instrument of the visual laboratory is a computer-controlled ADAGE visual display system that gives great flexibility in the type of display produced, with high spatial resolution, variable frame rates up to 120 fps, and color (if desired). The Adage is controlled by the VAX 11/750 and operates under the HIPS (the Human Information Processing Laboratory's Image Processing System) that affords users a powerful and versatile programming language. Displays that do not require full grey-scale throughout also can be produced on a custom made (Kropfl) interface operated from a PDP 11-23 computer, the display system that served during the first five years of USAF support. Recently, we have begun to add a modest auditory research facility to supplement the visual facilities.

PERSONNEL

Principle investigator, George Sperling, Professor of Psychology and director of the Human Information Processing Laboratory, 40% time (averaged over 12 months). The pronoun we is used in this proposal to refer to the PI in conjunction with one or more of the other investigators

19 APR 1988

88 5 16 115

and staff.

Full-time: Research Associate, Dr. Charles Chubb, who worked primarily on visual motion and on related mathematical issues.

Systems Programmer, Dr. Penelope Hall, who was visiting NYU from Cambridge, England, on limited-time appointment.

Part-time: Dr. Cathryn Downing, post-doctoral fellow supported by mainly by NEI who worked mostly on attentional issues related to the grant.

Dr. Barbara Dosher, a consultant who collaborated on projects in motion perception, visual memory and attention.

Karl Gegenfurtner, a graduate student, who worked on attention.

Roman Yangarber, a part-time student programmer.

Patrick Whelan, a part-time student administrative assistant who did accounting and related tasks.

The motion projects described in this report include the specification of low level motion detection systems which are both Fourier and NonFourier in kind, investigations in visual persistence in motion, higher level issues in structure from motion, cue integration and decision theory, and object recognition.

NonFourier Motion Perception.

Current mathematical or computational models of human motion perception, (e.g., Adelson & Bergen, 1985; Aloimonas & Brown, 1986; Fleet & Jepson, 1985; Heeger, 1986; Van Santen and Sperling, 1984, 1985; Watson and Ahumada, 1984, 1985) and models of machine vision that include motion perception as a front-end component (e.g., Aloimonas & Basu, 1986; Anandan, 1986; Anandan & Weiss, 1985; Bolles & Baker, 1985; Bulthoff & Mallot, 1987; Clocksin, 1980; Hildreth & Grzywacz, 1986; Shariat & Price, 1986; Subbarao, 1986a; Subbarao, 1986b; Waxman & Wohn, 1986) detect motion insofar as there is motion-related energy in the spatio-temporal Fourier transform of the display. Local Fourier analysis is central to van Santen & Sperling's elaborated Reichardt model, and these authors in the previous granting period proved and published the formal equivalence of the elaborated Reichardt model to Adelson & Bergen's and to Watson & Ahumada's models. Heeger's and the other recent developments are elaborations on this theme.

Indeed, the first and main motion project of the current grant was the specification of the Reichardt type models. (Background material on Reichardt detectors is provided below.) However, subsequently we (Chubb & Sperling) discovered that it was possible to create stimuli

that would be completely invisible to all of the models cited above and yet these stimuli yielded vigorous perceptions of movement. The investigation of such stimuli led to a whole class of experiments and theories relating to non-Fourier motion perception.

Decision tasks

Some interesting *decision* questions related to visual motion were been studied. Specifically, how does information from various kinds of motion detectors combine to determine the perception of motion (Dosher, Sperling & Wurst, 1985)? Previously, Burt and Sperling (1981) had found an additive rule for combining time Δt and distance Δx in ambiguous stroboscopic motion to arrive a measure of strength that predicted which of several alternative paths would be the one along which motion was perceived. In the current grant, two related issues were investigated. First, Dosher, Sperling, & Wurst (1986) investigated how stereopsis (the cue of binocular disparity) combined with the cue of proximity luminance covariance (the greater intensification of points nearer to the observer), in the perception of rotating 3D wire Necker cubes. These cubes are ambiguous, and can be perceived with equal probability to rotate clockwise or counterclockwise. Either cue alone was sufficient to bias the direction of perceived rotation overwhelmingly in a direction consistent with the cue. Together, either in synergy or in antagonism, the joint effect of the two cues was accurately accounted for by a simple model in which the individual effects of the two cues added linearly.

The simple additivity of the effect of cues is a natural prediction from energy (gradient descent) models of the sort proposed by Sperling (1970) and by Sperling, Pavel, Landy, Cohen, and Schwartz (1983). In this theory, evidence is weighed linearly, and the probability of perceiving one figural form (and the corresponding direction of rotation) rather than the other is proportional to the difference in the weight of evidence. In the computational vision literature, the combination of cues is a hot issue and under intensive investigation by Aloimonas (1986), Bulthoff and Mallott (1987), and others. There the issue is not so much the combination rule as how to use different cues in algorithms to reduce the number of alternative hypotheses about a visual scene.

Complex, Dynamic Visual Stimuli: ASL

Cue combination rules were studied by Riedl and Sperling (JOSA, 1988, in press) in a complex visual task -- the interpretation of American Sign Language (ASL). ASL was split into four spatial frequency bands of approximately equal intelligibility. Two combination questions were asked. First, how does noise in band i mask signal in band j to impair intelligibility and second, how does signal in band i combine with signal in band j to improve intelligibility. Riedl & Sperling discovered that the falloff with frequency separation of masking in ASL was quite comparable to masking with simpler stimuli. With respect to addition of signals, matters were quite different: addition seemed to be independent of the frequency distance between the component bands. Addition is complicated because many factors play a role: linear addition of amplitudes in the same band is efficient because it results in a nonlinear increase in power; linear addition of amplitudes in different bands results in only linear power increase but involves

different kinds of information. At low signal amplitudes, power considerations dominate; at high amplitudes, informational redundancy is critical.

In a related effort, Pavel, Sperling, Riedl & Vanderbeek (JOSA, 1987) studied the effect of signal to noise ratio in complex dynamic visual stimuli (ASL). They found that over an 8 to 1 range, stimulus contrast had no effect whatsoever on intelligibility of ASL, only the signal-to-noise ratio mattered. Further, they developed a simple computation framework in which the three most significant factors: signal-to-noise ratio, difficulty of an ASL sign, and competence of the viewer, were treated equivalently. That is, the difficulty of a particular sign can be represented as a noise source intrinsic to the sign with a power that is a proportional to signal power; competence of the viewer is represented as a factor that multiplies the signal-to-noise ratio. These simple relations offered a theoretically based, computationally efficient, detailed account of the data.

Structure from Motion. ***Overview.***

The eyes transmit only a two dimensional image. How does the brain use this information to reconstruct a representation of a three dimensional world? The profundity of the computation is especially evident in the Kinetic Depth Effect, the phenomenon in which a 2D representation of a 3D wire figure may be perceived as flat until the 3D figure begins to rotate, at which point the 2D image suddenly is perceived as a 3D object. In fact, 2D representations of wire figures, like all 2D figures, are inherently ambiguous when interpreted as 3D objects. This is especially important in understanding computer generated displays because many computer displays show outline objects, and even occasional misperceptions can be extremely costly.

In our research with rotating Necker cubes, the linear combination rules for sources of evidence were worked out with respect to stereopsis, proximity-luminance-covariance, perspective, context, hidden lines, and various other factors (see above, and Schwartz & Sperling; Dosher, Sperling & Wurst). The more recent work, a collaboration between Dr. Barbara Dosher, Dr. Michael Landy, Dr. Mark Perkins, and the PI, has dealt with the issue of what information is extracted from moving displays in order feed the structure from motion computation.

The first study dealt with the need for an objective task (Landy, Dosher, & Sperling, ARVO abstracts & Tech Report, 1987). It showed how various perceptual properties of KDE stimuli (such as perceived rigidity, perceived depth, and the perceived amount of fragmentation) could vary independently. The second study developed a new KDE task. In this task, the observer must discriminate between more than 50 quite similar alternative shapes that are depicted as a rotating surface sprinkled with randomly placed dots (Sperling, Landy, Dosher, Perkins, manuscript 1987). The third and fourth papers (Dosher, Landy, Sperling & Perkins, ARVO abstracts 1987, manuscripts in progress) show that various potentially useful sources of information (varying texture density, trackable features) can be eliminated without impairing performance and that optic flow information is both necessary and sufficient for computing structure from motion in these dot displays. In the early part of this proposal, our work on Fourier versus non-Fourier motion was described. The issue of whether the optic flow was computed by Fourier or non-

Fourier motion systems was addressed directly in the third and fourth studies. To a first approximation, it was found that only the Fourier motion system contributed significantly to the structure from motion computation (Dosher, Landy, Sperling & Perkins, 1987).

Six Studies of Structure from Motion.

Our first endeavor (B. Dosher, M. Landy, and G. Sperling) analyzed the use of such rating measures, and determined that there are actually several partially de-coupled aspects to such ratings. We showed observers displays generated by taking parallel or perspective projections of objects defined by dots either on the surface or in the volume of simple geometric figures like spheres or cylinders. We varied object, perspective, size, volume, and number of points. Observers were asked to rate the segmentation (whether the display represented one or several objects), depth (shape of trajectories in the depth plane), and rigidity, and showed that these were partially de-coupled aspects of displays. Incidentally, these experiments also showed that previously untested combinations of factors in more traditional KDE experiments, produced highly significant interactions in terms of their effect on the judged aspects of KDE.

As a result of these investigations, it became clear that a more profitable way to ask questions about KDE was to determine whether a particular display supported some level of performance on an objective task. An objective task was developed which targeted the critical component of KDE -- whether a 3D shape could be extracted or identified. A large lexicon (53 objects) of parametrically varied and easily named shapes was generated by sampling illuminated points from the shape's surface, and the lexicon was calibrated to verify a low guessing baseline (< 2%), and adequate identification performance (up to depth reversal) with modest point sampling and depth.

We discovered, to our surprise, that even in our complex, objective KDE task, subjects who extracted approximate 2D velocities in 6 locations could generate a correct response without a 3D perception of shape. Of course, subjects had to be instructed to this alternative computation, but it illustrates the significant problem of developing objective KDE measures. In this task, as with other previous work that used objective measures such as curvature identification (Todd, 1984), correct responses could, in principle, be "faked" -- answered on the basis of 2D correlates of the 3D motions--an alternative, non-KDE computation. Our task, however, requires a far more complicated scheme for alternative computation than prior tasks; it requires subjects to be specifically instructed in the alternative computation. For uninstructed subjects, the identification task seems appropriate for our work on the Input Problem.

The Input Problem. On the basis of extended examination of pilot displays which varied a large number of factors, we (B. Dosher and G. Sperling) became interested in a rather different kind of kinetic depth manipulation than has been common in the classic literature. Informal examination of displays which masked kinetic depth stimuli, reversed polarity, or interspersed grey background frames for extended durations led us to believe that adequate stimulation of classic low-level (elaborated Reichardt) motion detectors might be very important for supporting the extraction of structure from at least some kinds of KDE displays.

In formal experimentation on these factors the dependent measure is the proportion of correct identification of a shape from among the 53-element lexicon. (Formal experiments carried out with M. Landy and M. Perkins). Only one or two observers have provided data on the some of the manipulations listed below, so additional work remains to be done to yield a publishable product. The completion of this work is part of the current proposal, as well as several extensions.

The stimuli were illuminated points sampled randomly from the surface of a 3D shape from the lexicon. The lexicon is defined on three points of a triangle where the vertex points either up of down. These three points are either above, behind or on the back surface of a base-plane. These bumps and depressions generate a smooth form by spline techniques which connect the points in depth with the base frame and define the depth value for a 10,000 point grid. Initial calibration simply determined that sparse subsampling of 320 or even fewer points with a depth amplitude of .5 the side of the base plane yielded very high identification performance (85-95%) under standard conditions (sinusoidal rotation amplitude 25 deg, rotation period 30 frames, frame rate of 15 Hz, high contrast light on a dark background).

The main conditions concentrate on various manipulations of the motion cues to structure, so it was necessary to estimate and remove the contribution of the most obvious non-motion factor, namely density of illuminated points. Elimination of density cues to 3D shape required that sampled points be either removed or added to yield constant number in each small defined area of the display on each frame. The elimination of the density cue only marginally reduced identification accuracy; the density cue by itself generated slightly above chance performance (but far from motion-generated levels) for one subject and chance levels for several others. All subsequent displays eliminate the density cue.

A number of existing models of the extraction of a particular shape from KDE displays (Ullman, 1984; Landy, 1987) involve processes which develop depth for particular points over as long as a full rotation. Observation of pilot displays suggested that continued tracking of points in the display for three or more frames might not be necessary: informally, KDE appeared to be maintained with random rotation axis; initial data suggested that the cue-strength combinations were essentially the same for 2 frame displays as for subject-terminated displays (see the discussion of strength models below). In the identification paradigm, we verified that extended tracking of individual points or features, and measures of acceleration for individual points were not necessary for KDE. Multi-frame displays where every point is replaced by another random sample after two frames yielded quite good identification performance. Although performance was slightly lower than control levels, we suspect that the decrements are due to the introduction of scintillation (or temporal noise in correspondence) into the displays. Observers are known to be very sensitive to correspondence noise (Lappin et al, 1980).

Another set of manipulations were designed to be disruptive of low level motion analysis. The three main manipulations are interspersed grey (background) frames, polarity reversals, and generating displays with drift-balanced properties (see the definition of drift-balanced displays above). Extensive data were collected from one or two subjects for the polarity reversal data

along with a number of control conditions. Some data for the simple drift balanced displays are available. However, the boundary conditions of the effects need to be examined more carefully. We expect the work to be completed and written in 1988.

The polarity-reversal displays are generated so that sampled points may be either dark or light on a neutral grey background. (Prior conditions involve high contrast light points on a dark ground.) Each point reverses from light to dark (or vice versa) on alternating frames. Polarity reversal disrupts low level motion detectors, which may identify a polarity reversed stimulus as moving in the opposite direction. The polarity-reverse displays are very damaging to the extraction of depth structure from the displays. Control conditions compare the polarity-reversed displays with standard displays with contrast lowered sufficiently to yield equal performance on simple linear direction-of-motion judgments. Unlike these extremely low contrast non-reversing displays, initial data suggest that polarity reversed displays specifically suffer in motion segregation and structure from motion tasks. This suggests that strong evidence from the low-level motion systems is necessary, at least under some conditions, for effective 3D structure extraction.

Initial observations from the drift-balanced stimuli are equally suggestive. In order to examine drift balanced stimuli, it was necessary to extend our initial observations on point displays to displays with larger or thicker tokens. First we verified that structure identification is equally good when single (one pixel) sampled points are replaced with blobs several pixels in diameter, or when near sampled points are connected with short line segments (yielding a web of short lines over the shape to be recognized). In order to construct drift balanced stimuli, these larger tokens (blobs and lines) are replaced with random light/dark noise on a grey background. Either the light portion of these displays (replacing dark with grey background) or the dark portions yield quite good shape identification in KDE. When both appear together (under the careful calibration conditions necessary to drift-balancing -- where the grey is exactly midway between the light and dark), the depth structure collapses. This result also suggests the dependence of the structure-from-motion process on unambiguous output from low-level motion systems.

The strength problem. This section deals with how one or another percept comes to be seen in an ambiguous structure from motion display (where we assume the display is optimal or nearly optimal for generating KDE). Any 2D display of a 3D object admits of at least two percepts -- generally depth-reversed duals. With perspective transformation, these two duals possess one rigid and one rigid alternative (some displays are actually multi-stable, allowing more than one non-rigid percept). This is the problem, empirically, of what cues cause the observer to see a particular depth organization (either the rigid one or some non-rigid alternative). We completed and published (Dosher, Sperling, and Wurst, 1986) an analysis of how two particular cues combined to determine (statistically) which percept would dominate. This account varied stereopsis cues (favoring the rigid or non-rigid alternative) and luminance cues (favoring one or the other alternative). The cues could either agree, disagree, or could be neutral. We determined that these two cues are integrated particularly simply: the strength toward rigidity of each level of each cue (when scaled properly) simply add to determine the composite strength toward

interpreting a 2D display in the rigid mode.

As described above, the dependent measure in these studies was the proportion of trials in which the initial perceived rotation direction indicated a rigid percept. Additional experimentation showed that, not only did the additive cue model predict the proportion of trials, there was no evidence that conflicting cues (for example, stereo favoring rigid while luminance favored the nonrigid interpretation) were associated with longer response times for rotation direction. Further, response times could be very rapid; often response times are so short that only three frames of a rotating display (about 12 deg of rotation) could have been seen prior to the response. Indeed, preliminary data from one subject can be described quite well by a single set of strength and parameter estimates for conditions where only 2 frames, 3 frames, 4 frames or unlimited frames were shown. These data suggest that, at least for this subject and for familiar wire objects, essentially the same perceptual processes, with nearly identical strengths and weights, underlie performance for both unlimited time displays and displays with as few as 2 frames. This is related to the findings (described above) for multi-frame displays with features restricted to 2-frame lifetimes.

The shape computation problem. Given essentially ambiguous displays, such as those used in the cue-strength experiments above, the major computational problem has been defining a model which (i) computes a depth organization from a 2D display; (ii) predicts not just the rigid alternative, but the non-rigid alternative in ambiguous displays, and (iii) provides a rigidity metric which predicts the perceived "rubberiness" in a selected depth organization. Two theoretical approaches to the structure-from-motion problem propose either computations based on low-level velocity or motion information (e.g., Koenderink and van Doorn, 1986; Hoffman, 1982; Hildreth & Grzywacz, 1986) or those based on tracking of object features (Ullman, 1984; Landy, 1987). Both methods restrict the possible solutions by a rigidity assumption (extract the rigid object), or some related assumption.

In one project, Dosher and Sperling analyzed the class of 3D distance rigidity-based models of which Ullman (1984) and Landy (1987) are examples. (The same properties would hold for any model which defined its fidelity criterion on a 3D distance space, for example models desiring constant 3D velocity.) These rigidity-based models essentially perform a parameter search to find the best depth coordinates for features in the image. The search is designed to minimize an error function (nonrigidity), or maximize a fidelity criterion (rigidity). The proposed criteria are monotonic functions of the changes in the inter-feature (interpoint) distances in the extracted (estimated) 3D object. It is known that such models can find either the true object, or its depth reversed dual (under some boundary restrictions) under parallel projection. In most instances, both the Ullman and the Landy model find a correct solution rather slowly, requiring half or more of a full rotation and numerous frames to converge.

We were able to show that a necessary pre-condition of correct performance of these models was a match between the type of perspective assumed by the model or algorithm and that used to generate the 2D image. The published algorithms have assumed parallel projection and have been tested on parallel images. (The authors have implied that the issue of perspective

could be finessed, locally parallel solutions being combined by some unspecified higher process.) When the parallel algorithms operate on perspective images, two depth-reversed duals are extracted which are exactly identical, and so are identical with respect to non-rigidity. In contrast human observers extract one object which is seen as rigid, and a different (more than depth reversed) object which is seen as quite rubbery over rotation. We were further able to show that such distance-based rigidity algorithms could perform like the human observers only if the correct perspective transformation were known.

These findings were stated as symmetry properties of the energy surfaces in the structure from motion problem, and we computed some highly simplified energy surfaces under special assumptions to illustrate the point.

Visual Persistence in Motion Perception.

When continuous motion of a point is represented in a series of frames (as in movies, TV, and most computer displays), the smooth continuous movement is approximated by a series of sampled points. When the point moves a considerable distance from frame to frame, the observer will probably see multiple images of the point. This is because the image of the point in one frame persists in the visual system while subsequent points are being displayed. In fact, Farrell, Pavel, and Sperling (1983) developed an improved psychophysical method based on the number of visible points for estimating the duration of visual persistence. A paper describing new data and a theory for estimating the degree of persistence as a function of the distance between adjacent points is in preparation.

Visual Attentive Processes

These projects dealt with the ability of humans to process information arriving simultaneously at different locations in the visual field and to coordinate concurrent visual and auditory inputs. A major theme was the measurement of the time taken by an observer to shift attention from one area of the visual field to another, the consequences of this shift for information processing at both locations. In a theoretically related (but practically quite different) project memorability in auditory memory was investigated as a function of item familiarity, item length, and interitem confusability. A common goal of all projects was the description of the human abilities and limitations in the allocation of mental processing resources, and correspondingly, the theoretical derivation of visual and auditory stimulus codes that take optimum advantage of human abilities. This knowledge will eventually aid in the design of better and more reliably interpreted displays, in better training procedures, and in better assessments of individual differences in performance potential.

The work on attention was reviewed in several lengthy publications (Sperling, 1984, Unified Theory of Attention and Signal Detection; Reeves and Sperling, Pschol Rev., 1985, Gating Theory of Attention; Sperling & Dosher, AF Hbk, 1986, Strategy and Optimization in Human Information Processing), and one brief report Weichselgartner & Sperling, Science, 1987, Dynamics of Automatic and Controlled Attention) so there is no need to dwell on it here. We

consider progress in three areas during the last year of the grant.

Dual processes in attention shifts.

An essential component of visual attention is attentional gating, the process whereby some incoming information is selected for further analysis or for memorization, while other information is ignored or attenuated and lost. Normally, eye movements and attentional shifts are tightly coupled, but we are concerned with attentional processes that can occur while the eyes are stationary. Early in the current grant period, in the investigation of the dynamics of attentional gating, we were measuring the time course of attention with two kinds of stimuli: a cue to begin attending or to shift attention and a stimulus to be attended. We noticed curious bimodal distributions of "attention shift times" that suggested that we were observing not merely a single act of attention but two consecutive, partially overlapping acts. During the course of investigating these phenomena, we learned how to attain separate and almost independent control of the time course of each attentional processes. Specifically, in the grant period, Weichselgartner and Sperling (Weichselgartner & Sperling, 1986a, 1986b, 1987) studied attentional episodes in a task in which subjects monitored a stream of digits being presented at a single spatial location at a rate of 10 digits per sec. Their task was to remember the digit that appeared with a square surrounding it and the three digits that subsequently followed. These studies revealed two attentional processes, one apparently automatic process that was engaged immediately upon the appearance of the outline square, and another apparently more effortful and controlled process that was not engaged for some 200 to 300 msec following the appearance of the square. The process first engaged by the square was not only very quick, but was also quite potent, allowing subjects to report the digit appearing within the square on virtually 100% of the trials. Digits occurring 100 or 200 msec following the square were sometimes brought in within this "first glimpse" but with much lower probability. The first process is a quick, effortless, automatic process triggered by target detection, that records the cue to begin attending and its neighboring events. The second is a slower, effortful, controlled process that records the stimuli to be attended, and whose latency depends on practice and task difficulty. A report of this research was published in Science (1987).

The Time Course of Iconic Memory

Low-level, relatively unprocessed visual memory (here called VM1; previously called iconic memory [Neisser, 1967], or visual information storage [Sperling, 1959]) has been studied for many years. During the previous grant period we developed a new method of measuring the entire time course of the perceived rise and decay of visual persistence by means of comparisons to standard-intensity displays (Weichselgartner & Sperling, 1985).

Levels of Visual Processing.

Sperling and Jane Kaufman (Kaufman, 1977) provided evidence for low-level visual persistence which is distinct from iconic memory, or VM1. This persistence, VM2, was revealed in the context of a repetition-detection paradigm. VM2 is a *form* (e.g., type-font) and *location* dependent visual representation that underlies the detection of repetitions (with at most one intervening item) of digits presented in rapid succession at the same location in space. VM2 is distinct from iconic memory (VM1) because it survives masking by a high-intensity pattern mask. VM2 is distinct from more abstract visual memories (VM3) because it depends on the particular typefont and precise location of a character.

In the current grant period, this research was extended by exploratory investigations that determined that VM2 or VM3 were not sensitive to eye of origin of information, nor were they sensitive to location of that information (when only two different locations were involved). Attempts to parameterize this non-iconic short-term visual memory in terms of a signal detection model that concurrently treats accuracy and confidence were quite successful, and cast doubts on the earlier interpretation of the data as resulting from two memory processes. S. A. Wurst made a good start on a Ph.D. thesis on this topic during the grant period, but the completion and publication will not occur before late 1988 or 1989.

Auditory Memory Models, Optimal Codes.

Overview. Some of the principles developed in the study and description of visual memories are also applicable to the study of auditory memory. For example, in the structural description of a memory, what factors optimize it for recognition (like visual memory) and what factors optimize it for recall (like auditory memory). Given a structural description of a memory it should be possible to predict the optimal code for any particular performance, given the memory constraints. For example, it had been repeatedly observed that short-term auditory memory for digits consistently was better than auditory memory for letter lists, even when the number of alternatives from which the list items was drawn was the same in both lists.

In the current grant period, a theoretically (phonemically non-confusable) optimal set of 9 letters of the alphabet was derived. When short term memory for lists composed of elements from this set was tested, it was very nearly equal to memory for digits. The much improved short-term memory performance with the special letter set demonstrates the power of the phonemically-based predictions. However, the limits of phonemic predictions were demonstrated earlier by showing that less familiar items are remembered less well than more familiar items, even when they are phonemically matched. In the current grant period, several protracted attempts to train subjects with small sets of unfamiliar items to improve the short-term memorability of these items have been unsuccessful. In a formal multi-factor experiment, it was determined that the memory effects of phonemic structure and of familiarity were simply additive. At this time, we do not know what makes some sets of items, such as digits and letters, more memorable in STM than sets of equally short common words. These issues must be resolved in subsequent work.

Codes that Optimize Memorability. To design a code that will enable messages in that code to be remembered optimally involves many considerations that are similar to those in designing an optimally intelligible code. Research conducted in this laboratory (G. Sperling and collaborators) is described below. The starting fact is that spoken strings of digits are remembered better than strings of letters or strings of words. The initial question asked is: To what extent is the particular phonemic structure of digits responsible for our good memory for lists of digits? To learn more about the role of phonemic structure, *per se* a set of pseudodigits that was phonetically matched to the digits was produced as follows. The phonemes of English were divided into mutually exclusive pairs, so as to maximize the similarity of the phonemes within each pair, i.e., (m,n), (r,l), (s,f), (ey,ay), (i,u), etc. To produce a set of pseudodigits that was phonemically matched to digits, each phoneme of digits "oh, one, two, ..., nine" was replaced with its pair mate, yielding "ouw, yim, key, plu, sal, says, futz, fizim, ike, mame."

Several tokens of each digit and each pseudodigit were recorded by a male speaker with normal American speech. One sample of each type was selected from the recording and this token was used for all subsequent lists.

A list of length n, $6 < n < 11$, was prepared by selecting randomly, with replacement, n times from either the digit set or from the pseudodigit set. Items from the two sets were never mixed within a list. The stimulus items occupied 600 msec; there was a 70 msec pause added after every third item to produce temporal groups of three items, plus zero, one, or two remainder items after the last group.

To measure short-term memory for these digits and pseudodigits, tape recorded lists of items were played to subjects through earphones. The subjects were instructed to repeat aloud each list immediately upon its termination. Subsequently, responses were recorded and scored for the number of items reported correctly in their correct serial positions. Blocks of ten trials were conducted at each of five list lengths that maximized each subject's score. (Different sets of lengths were used for digits and pseudodigits.) A daily session consisted of 20 blocks (200 trials) conducted in a counterbalanced order. Three subjects served for eight to ten sessions.

In addition to the STM tests, subjects were given recitation training to determine whether it would improve their recall performance. They were required to learn to produce two consecutive errorless recitations of the following digit (and corresponding pseudodigit) lists both forward and backward: 0,1,2,3,4,5,6,7,8,9; 0,2,4,6,8,1,3,5,7,9.

Results (1) The recitation training had no effect upon performance. (2) For every subject and every list length, recall was better for digits than pseudodigits. The recall deficit for pseudodigits, averaged over all subjects and over comparable lists, was 1.17 items (per recalled list). After the initial session, this deficit showed no tendency to diminish.

Subjects achieved the highest scores with lists of length 8 and 9 with digits and with lists of length 7 and 8 for pseudodigits. Comparing recall at the optimal list length for each type of stimulus material yielded a mean pseudodigit deficit of 1.45 digits.

The conclusion is that digits have a substantial recall advantage over pseudodigits. The digit/pseudodigit advantage is quite resistant to considerable practice (about a thousand trials with each type of list) and to modest amounts of recitation training. Insofar as we can exclude as yet undiscovered artifacts of procedure as being responsible for these results, the advantage of digits must lie in their greater familiarity, where the dimensions of familiarity now need to be explored further. No model of short-term memory that deals only with the phonetic structure of the elements to be recalled can be adequate to account for these results.

Recall factors: Familiarity, phonemic structure, and acoustic confusability. The experiment reported above showed that familiarity (the prior exposure to digits in our culture) accounted for their superior memorability over pseudodigits that were matched in phonemic structure and in acoustic confusability. What is the relation of recall for letters (which average 1.9 phonemes per letter) to recall for digits (which average 3.1 phonemes)? In fact, a long history of observation has indicated that letter recall is inferior to digit recall. Is this due to lack of acquired familiarity with letter strings, to the different phonemic structure, or to acoustic confusability of letters?

Our experiments with lists of letters showed that artificially constructed pseudoletters suffered as much in recall relative to letters as pseudodigit suffered relative to digits. This excludes familiarity. Apparently, years of practice with letter strings also makes letters memorable. Previous research (by Conrad, Sperling, and others) showed that confusable letter lists (b,c,d,t,v etc or n,m,l,s,f etc) were remembered much less well than nonconfusable lists. One outcome of our recent research has been that, by constructing the list of optimally distinct letters for the letter/pseudoletter experiments, for the first time lists of letters were remembered as well as lists of digits. Thus acoustic confusability seems to account for the usual letter/digit recall difference.

The bottom line is that acoustic confusability (since the 1960s) and now familiarity have been shown to be important determinants of short-term memory for spoken lists. Phonemic structure (which was the the basis of Sperling's (1968) model that accounted for short-term memory for letters and for the acoustic confusability deficit) appears to have little influence, at least for items of two or three phonemes. Obviously, for very long items such as those that are generated by codes that optimize intelligibility, memorability will be compromised.

Publications

1984 Sperling, G. A unified theory of attention and signal detection. In R. Parasuraman and D. R. Davies (Eds.), *Varieties of Attention*. New York, N. Y.: Academic Press, 1984. Pp. 103-181.

1984 van Santen, Jan P. H., and George Sperling. Temporal covariance model of human motion perception. *Journal of the Optical Society of America A: Optics and Image Science*, 1984, *1*, 451-473.

1984 Van Santen, Jan P. H., and George Sperling. Applications of a Reichardt-Type Model to Two-Frame Motion. *Investigative Ophthalmology and Visual Science, ARVO Supplement*, 1984, *25*, 14. (Abstract)

1984 Reeves, Adam, and George Sperling. Visual temporal order perception. *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1984, *25*, 69. (Abstract)

1984 Weichselgartner, Erich, and George Sperling. Psychophysical method for the continuous measurement of visible persistence. *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1984, *25*, 70. (Abstract)

1984 Landy, Michael S., and George Sperling. Image Processing in Visual Psychophysical Research. *Investigative Ophthalmology and Visual Science*, ARVO Supplement, 1984, *25*, 72. (Abstract)

1984 Landy, Michael S., Yoav Cohen, and George Sperling, HIPS:† Image processing under Unix. Software and applications. *Behavior Research Methods and Instrumentation*, 1984, *16*, 199-216.
†HIPS is the Human Information Processing Laboratory's Image Processing System.

1984 Sperling, George. Image processing studies of American Sign Language. *Journal of Psycholinguistic Research*, 1984, *13*, 494. (Abstract)

1984 George Sperling, David H. Parish, M. Pavel, and Douglas H. Desaulniers. Auditory list recall: Phonemic structure, acoustic confusability, and familiarity. *Bulletin of the Psychonomic Society*, 1984, *22*, No. 4, Page 292. (Abstract)

1984 Barbara Anne Dosher, George Sperling, and Steven Wurst. Stereopsis versus proximity-luminance covariance as determinants of perceived three-dimensional structure. *Journal of the Optical Society of America A: Optics and Image Science*, 1984, *1*, 1254. (Abstract)

1984 Landy, Michael S., Yoav Cohen, and George Sperling. HIPS†: A Unix-based image processing system. *Computer Vision, Graphics, and Image Processing*, 1984, *25*, 331-347.
†HIPS is the Human Information Processing Laboratory's Image Processing System.

1985 van Santen, Jan P. H., and George Sperling. Elaborated Reichardt detectors. *Journal of the Optical Society of America A: Optics and Image Science*, 1985, *2*, 300-321.

1985 Sperling, George, J. P. H. van Santen, and Peter J. Burt. Three paradigms and three theories of stroboscopic motion perception. *Spatial Vision*, 1985, *1*, 47-56.

1985 Erich Weichselgartner, George Sperling, and Adam Reeves. Effects of Concurrent Tasks, Distance, and Visual Obstacles on Shifts of Visual Attention. *Proceedings of the Eastern Psychological Association*, 1985, *56*, 72. (Abstract)

1985 George Sperling and David H. Parish, Forest-in-the-Trees Illusions. *Investigative Ophthalmology and Visual Science, ARVO Supplement*, 1985, *26*, 285. (Abstract)

1985 George Sperling, Michael Landy, Yoav Cohen, and M. Pavel. Intelligible Encoding of ASL Image Sequences at Extremely Low Information Rates. *Computer Vision, Graphics, and Image Processing*, 1985, *31*, 335-391.

1985 Weichselgartner, Erich, and George Sperling. Continuous measurement of visible persistence. *Journal of Experimental Psychology: Human Perception and Performance*, 1985, *11*, 711-725.

1985 Landy, Michael S., Barbara A. Dosher, and George Sperling. Assessing kinetic depth in multi-dot displays. *Bulletin of the Psychonomic Society*, 1985, *23*, No. 4, Page 295. (Abstract)

1986 Sperling, George. A signal-to-noise theory of the effects of luminance on picture memory: Comment on G. Loftus. *Journal of Experiental Psychology: General*, 1986, *115*, 189-192.

1986 Dosher, Barbara A., Sperling, G., & Wurst, S. A. Tradeoffs between stereopsis and proximity luminance covariance. *Vision Research*, 1986, *26*, 973-990.

1986 Reeves, Adam, and George Sperling. Attention gating in short-term visual memory. *Psychological Review*, *93*, 180-206.

1986 Weichselgartner, E. and Sperling, G. Two processes in visual attention. *Proceedings of the Eastern Psychological Association*, 1986, *57*, 38. (Abstract)

1986 Sperling, George, and Barbara Anne Dosher. Strategy and optimization in human information processing. In K. Boff, L. Kaufman, and J. Thomas (Eds.), *Handbook of Perception and Performance. Vol. 1.* New York, NY: Wiley, 1986. Pp. 2-1 to 2-65.

1986 Weichselgartner, Erich, and George Sperling. Gleichzeitige Erfassung des Zeitverlaufs Zweier Visueller Aufmerksamkeitsprozesse. *Bericht ueber den 35. Kongress der Deutschen Gesellschaft fuer Psycholgie in Heidelberg 1986. Band 1. Kurzfassung.* Goettingen: Verlag fuer Psycholgie, 1986. P. 99.

Weichselgartner, Erich, and George Sperling. Estimation of the time course of two concurrent attentional processes. *Proceedings of the 35-th Congress of the German Psychological Society. Vol. 1.* Summaries. Goettingen: Verlag fuer Psycholgie, 1986. Page 99. (Abstract)

1986 Sperling, George. and Weichselgartner, Erich. Measuring the time course of automatic and controlled attention. *Bulletin of the Psychonomic Society*, 1986, *24*, No. 5, Page 347. (Abstract)

1987 George Sperling and Barbara Anne Dosher, Predicting rigid and nonrigid perceptions. *Investigative Ophthalmology and Visual Science, 28, No. 3, ARVO Supplement*, 1987, 362. (Abstract)

1987 Landy, Michael S., George Sperling, Barbara Anne Dosher and Mark Perkins, From what kind of motions can structure be inferred? *Investigative Ophthalmology and Visual Science, 28, No. 3, 233.* (Abstract)

1987 Chubb, Charles, and George Sperling, Drift-balanced random stimuli: A general basis for studying nonFourier motion perception. *Investigative Ophthalmology and Visual Science, 28, No. 3, ARVO Supplement*, 1987, 233. (Abstract)

1987 Parish, David H. and George Sperling, Object spatial frequency, not retinal spatial frequency, determines identification efficiency. *Investigative Ophthalmology and Visual Science, 28, No. 3, ARVO Supplement*, 1987, 359. (Abstract)

1987 Weichselgartner, E., and George Sperling. Dynamics of automatic and controlled visual attention. *Science*, 1987, *238*, 778-780.

1987 Sperling, George. The magical number seven: Information processing then and now. In William Hirst (Ed), *Giving Birth to Cognitive Science: A Festschrift for George A. Miller*. Cambridge, UK: Cambridge University Press, 1987 (In press).

1987 Michael S. Landy, Barbara A. Dosher, George Sperling, and Mark E. Perkins. Perception of complex shape from optic flow. *Journal of the Optical Society of America A: Optics and Image Science*, 1987, *4*, *No. 13*, P95. (Abstract)

1987 Pavel, M., George Sperling, Thomas R. Riedl, and August Vanderbeek. The limits of visual communication: The effects of signal-to-noise ratio on the intelligibility of American Sign Language. *Journal of the Optical Society of America - A*, 1987, *4*. 2355-2365.

1988 Riedl, Thomas R. and George Sperling. Spatial frequency bands in complex visual stimuli: American Sign Language. *Journal of the Optical Society of America - A*. 1988, *5*. (In press.)

Technical Reports, Papers Under Submission for Publication

1985 Weichselgartner, E. and George Sperling, G. The effects of distance, concurrent tasks, and visual obstacles on shifts of visual attention. *Mathematical Studies in Perception and Cognition, 85-3 (Rev)*, New York University, Department of Psychology, 1985. Pp. 36, Figs. 9.

1987 Sperling, George, Michael S. Landy, Barbara A. Dosher, and Mark E. Perkins. The kinetic depth effect and the identification of shape. *Mathematical Studies in Perception and Cognition, 87-1 (Rev)*, New York University, Department of Psychology, 1987. Pp. 30, Figs. 3.

1987 Parish, David H. and George Sperling, Object spatial frequency, retinal spatial frequency, and the efficiency of identification. *Mathematical Studies in Perception and Cognition, 87-8*, New York University, Department of Psychology, 1987. Pp. 30, Figs. 8.

1987 Barbara A. Dosher, Landy, Michael S., and George Sperling. Ratings of kinetic depth in multi-dot displays. *Mathematical Studies in Perception and Cognition, 87-3 (Rev)*, New York University, Department of Psychology, 1987. Pp. 24, Figs. 5.

1987 Chubb, Charles, and George Sperling. Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Mathematical Studies in Perception and Cognition, 87-7*, New York University, Department of Psychology, 1987. Pp. 64, Figs. 11.

1987 Farrell, Joyce E., M. Pavel, and George Sperling. The visible persistence of stimuli in stroboscopic motion. *Mathematical Studies in Perception and Cognition, 87-9*, New York University, Department of Psychology, 1987. Pp. 20, Figs. 9.

END
DATE
FILMED
7-88
DTIC